Speech-Based Query Assistant Using Machine Learning and Natural Language Processing

Maradana Purnima[1], Umamaheswararao Mogili[2*], T. Syam Prasad[3], B. Uday Bhaskar Sai[4], V. Santhosh Kumar[5], P. Rohit Kumar[6], G. Jhansi[7]

[1,2*]*Assistant Professor, Department of Computer Science and Engineering, Avanthi's St Theressa Institute of Engineering and Technology, Garividi, Andhra Pradesh, India*

[3, 4,5,6,7]*B.Tech, Department of Computer Science and Engineering, Avanthi's St Theressa Institute of Engineering and Technology, Garividi, Andhra Pradesh, India*

Abstract: Voice Query Assistants (VQAs) are becoming an important part of modern technology, allowing people to interact with computers using voice commands instead of typing. These assistants make it easier and faster for users to perform tasks and access information. The main goal of this project is to design and develop a Voice Query Assistant that can understand user voice commands and respond accordingly. In this project, a voice assistant named MEERA is developed using the Python programming language. The system uses technologies such as Automatic Speech Recognition (ASR) to convert speech into text and Text-to-Speech (TTS) to respond back to the user with voice output. It also uses Natural Language Processing (NLP) techniques to understand the user's query and provide suitable responses. A simple and user-friendly Graphical User Interface (GUI) is created using Tkinter so that users can easily interact with the assistant. The assistant is capable of performing several useful tasks such as searching information on Wikipedia, opening Google or YouTube, playing videos, checking the weather, telling the current time, opening system applications like Notepad or Calculator, checking system conditions like CPU usage and battery percentage, and even sending emails. By using voice commands, users can interact with the computer in a more natural and convenient way. This project shows how voice-based systems can improve the user experience by making technology more accessible and easier to use. It also demonstrates the practical use of voice recognition and artificial intelligence in everyday applications. Although voice assistants provide many advantages, challenges such as background noise, speech recognition accuracy, and internet dependency still exist. Despite these limitations, voice assistants have great potential and will continue to play an important role in future human–computer interactions.

Keywords: Voice Assistant, Speech Recognition, Natural Language Processing, Python, Human– Computer Interaction, Voice-Based Systems.

I. INTRODUCTION

The very first voice activated product was released in 1922 as Radio Rex. This toy was very simple, wherein a toy dog would stay inside a dog house until the user exclaimed its name, "Rex" at which point it would jump out of the house. This was all done by an electro magnet tuned to the frequency similar to the vowel found in the word Rex, and predated modern computers by over 20 years. In the 21st century, human interaction is being replaced by automation very quickly. One of the main reasons for this change is performance. There's a drastic change in technology rather than advancement. In today's world, we train our machines to do the tasks by themselves sort to think like humans using technologies like Machine Learning, Neural Networks, etc. Now in the current era, we can talk to our machines with the help of virtual assistants. Virtual assistants are software programs that help you ease your day-to-day tasks, such as showing weather reports, giving daily news, searching the internet etc. They can take commands by voice. Voice-based intelligent assistants need an invoking word or wake word to activate the listener, followed by the command. We have so many virtual assistants, such as Apple's Siri, Amazon's Alexa and Microsoft's Cortana and Amazon's Alexa and this has been an inspiration for us to do this as a project. This system is designed to be used efficiently on desktops. Voice assistants are programs on digital devices that listen and respond to verbal commands. A user can

say, "What's the weather?" and the voice assistant will answer with the weather report for that day and location.

II. LITERATURE SURVEY

Voice assistants and voice-based search systems have become an important area of research due to the rapid growth of speech recognition technologies. Several studies have focused on improving query understanding, speech recognition, and information retrieval through voice interfaces. Query suggestions are widely used to clarify the intent of incomplete or underspecified queries in voice-based search environments [1, 2]. Secure Data Storage and Sharing in Multi-Cloud Environment In the cloud storage is also described to store the predicted data in a secured way [3-8]. SQL has also been explored as a backend technology for natural language database programming, enabling users to interact with databases through conversational queries. The increasing use of voice interactions has changed the way users communicate with search systems. Users often reformulate their queries when the system fails to recognize speech correctly or when the returned results do not match their expectations [9-10]. The popularity of mobile search and advancements in speech recognition has further enabled users to speak their queries instead of typing them. Virtual Assistants (VAs) such as Apple Siri, Amazon Alexa, Microsoft Cortana, and Google Assistant have emerged as major platforms for spoken information retrieval and task automation [11-14]. These assistants allow users to perform tasks such as searching for information, controlling smart devices, managing schedules, and accessing services through voice commands. Voice Assistant Systems are also being integrated into healthcare environments to assist patients and caregivers by enabling voice-based interaction with medical systems. Despite their widespread adoption, voice assistants face several challenges such as linguistic variations, speech pattern variability, background noise, and acoustic disturbances, which often lead to misinterpretation of user queries. To improve the quality of responses and enhance user experience, query reformulation techniques and speech recognition improvements are being widely studied. Voice assistants are now widely used across various platforms including smartphones, smart speakers, and home automation systems. Their ability to support natural language interaction and context-aware conversations has significantly transformed the way users interact with technology [15-17]. These advancements highlight the importance of improving speech recognition accuracy, query processing techniques, and intelligent response generation in modern voice assistant systems. Some of the sample artificial intelligence, machine learning and deep learning models for prediction for fire detection are described in details [1-8].

III. METHODOLOGY

3.1. Problem Definition and Scope
When developing a voice assistant, the first step is to clearly understand the purpose of the system. The goal of the voice assistant should be defined based on the problem it is intended to solve. For example, it may be designed to help with customer service, control smart home devices, or provide quick information such as weather updates, news, or general knowledge. It is also important to identify the target users who will interact with the system. These users may include tech-savvy individuals, the general public, or people with visual impairments who rely on voice-based interfaces for accessibility. Another important factor is deciding the domain of the assistant. Some voice assistants are designed to answer a wide range of general questions, while others focus on a specific area such as answering frequently asked questions, providing weather information, or supporting a particular service.

3.2. System Architecture Design
After defining the purpose of the assistant, the next step is designing the overall system architecture. A voice assistant typically works through a series of connected components that handle different tasks. First, the system needs to receive the user's voice as input and convert it into text using speech recognition. Once the speech is converted into text, the processing stage begins. In this stage, the system tries to understand the meaning of the user's request using Natural Language Understanding techniques. The dialog management component then decides how the assistant should respond, and it may also retrieve information from a knowledge base or external APIs. Finally, the generated response is converted back into speech so that the user can hear the assistant's reply.

3.3. Voice Input Processing

Voice input processing is responsible for capturing and understanding the user's spoken commands. This process uses Automatic Speech Recognition (ASR) technology to convert spoken words into written text. Several tools and services can be used for this task, including Google Speech-to-Text, Microsoft Azure Speech, Mozilla Deep Speech, and Whisper. While selecting a speech recognition system, it is important to consider factors such as support for different languages and accents, the ability to work accurately in noisy environments, and whether the system can process speech in real time or only after the audio is recorded.

3.4. Natural Language Understanding

Once the speech is converted into text, the system needs to understand what the user actually means. This is done through Natural Language Understanding. In this step, the assistant identifies the user's intent, which represents the action or information the user is requesting. At the same time, it extracts important details from the sentence, such as names, locations, or dates. These details are known as entities. Various libraries and frameworks such as Rasa NLU, spaCy, BERT, Dialog flow, and LUIS can be used to perform this task. In some cases, especially when the assistant is designed for a specific field or domain, the model may need to be trained with custom data to improve accuracy.

3.5. Dialogue Management

Dialogue management controls how the conversation flows between the user and the voice assistant. In simple systems, each command is handled separately, and the assistant responds without remembering previous interactions. This type of system is known as stateless. However, more advanced assistants are stateful, meaning they can remember previous parts of the conversation and maintain context across multiple interactions. This allows the assistant to have more natural and meaningful conversations with users. Tools such as Rasa Core, Microsoft Bot Framework, and Dialog flow CX can be used to manage and design these conversational flows.

3.6. Backend and Knowledge Base Integration

To provide useful responses, the voice assistant often connects to a backend system or knowledge base. In some cases, the assistant simply provides predefined responses for common questions, such as FAQs. In other situations, it may retrieve information dynamically by connecting to databases or external APIs. For example, the assistant may use a weather API to provide current weather information or connect to a news service to deliver the latest updates. It is also important to design a fallback mechanism that allows the system to respond appropriately when it cannot understand the user's query or when the request falls outside its capabilities.

3.7. Voice Output using Text-to-Speech

After the system generates a response, it must convert the text into spoken language so the user can hear it. This process is handled by Text-to-Speech technology. Popular TTS engines include Google Cloud TTS, Amazon Polly, IBM Watson TTS, and Coqui TTS. When designing the voice output, developers often consider factors such as voice tone, gender, speaking speed, and expressiveness. These elements help make the assistant sound more natural and pleasant for users during interaction.

3.8. Evaluation and Testing

Testing and evaluation are essential to ensure that the voice assistant works effectively. Different metrics are used to measure the performance of each component. For example, Word Error Rate is used to evaluate how accurately the speech recognition system converts voice into text. Intent accuracy measures how well the system identifies the user's intention. Response accuracy evaluates whether the assistant provides correct and helpful answers. In addition to technical performance, the overall user experience must also be considered. This includes factors such as response time, ease of interaction, and user satisfaction. Testing methods may include manual testing, user feedback surveys, and analytical tools such as confusion matrices.

3.9. Iterative Improvement

A voice assistant is not a system that is built once and left unchanged. Instead, it requires continuous improvement based on real user interactions. Developers can collect user feedback to understand the strengths and weaknesses of the system. The

training data used for speech recognition and natural language understanding can be expanded and improved over time. The conversation logic can also be refined to make responses clearer and more helpful. Techniques such as A/B testing can be used to experiment with different voice styles and response strategies to determine which ones provide the best user experience.

3.10. Deployment and Monitoring

Once the voice assistant is ready, it can be deployed on different platforms such as websites, mobile applications, or smart devices. After deployment, it is important to continuously monitor the system to ensure that it performs well in real-world conditions. Monitoring may involve tracking user queries, analyzing system errors or crashes, and studying usage patterns to understand how people interact with the assistant. At the same time, developers must pay close attention to privacy and ethical considerations. This includes following data protection regulations, ensuring that user data is anonymized when necessary, and obtaining proper consent before collecting or processing voice data.

IV.   RESULTS & DISCUSSION

The MEERA Voice Queries Assistant was successfully developed and tested. The system performs voice-based interaction through a graphical user interface. An entering the user's name, the assistant greets the user based on the current me and begins listening for voice commands. Sample Inputs And Output:

1. User Name Input
Input:
User enters name: SYAM Output (Voice Response):
"Hello SYAM"
"Good morning SYAM" (Based on me) "I am Meera. How may I assist you?"

2. Open Notepad
Voice   Input: "Open notepad" Output:
Assistant says: "Opening Notepad for you sir" System opens Notepad application.

3. Open YouTube
Voice Input: "Open YouTube"
Assistant Response: "What do you want to play on YouTube sir?"

Voice Input: "Python tutorial"
Output: Assistant plays Python tutorial video on YouTube in browser.

4. Wikipedia Search
Voice Input: "Wikipedia"
Assistant Response: "What do you want to search on Wikipedia sir?"
Voice Input: "Artificial Intelligence"
 Output: Assistant reads summary of Artificial Intelligence. Summary is also printed in terminal.

5. Weather Report
Voice Input: "Weather"
Assistant Response: "Tell me the name of your city"
Input (Keyboard): Hyderabad
Output: "The current temperature is 30 Fahrenheit, but it feels like 32 Fahrenheit."
Assistant says: "Also, the weather report talks about Clear sky." Weather details printed on screen.

6. Tell Time
Voice Input: "What is the time now
Output: "The me is 10 hours 30 minutes and 15 seconds." Time also printed in terminal.

7. System Condi on
Voice   Input: "System condition
"Output": "CPU is at 45 percent."
"Boss our system has 80 percent battery."
"Boss we have enough charging to continue."

8. Calculator
Voice   Input: "Open calculator
" Output":
Assistant says: "Opening your calculator" Calculator application opens.

9. Send Mail
Voice Input:
"Send mail"
Assistant Response: "Enter receiver's mail and Message" Input:
Receiver: example@gmail.com Message: Hello, this is a test message.
Output: Assistant says: "Message sent" Email successfully delivered.
10. Opening map
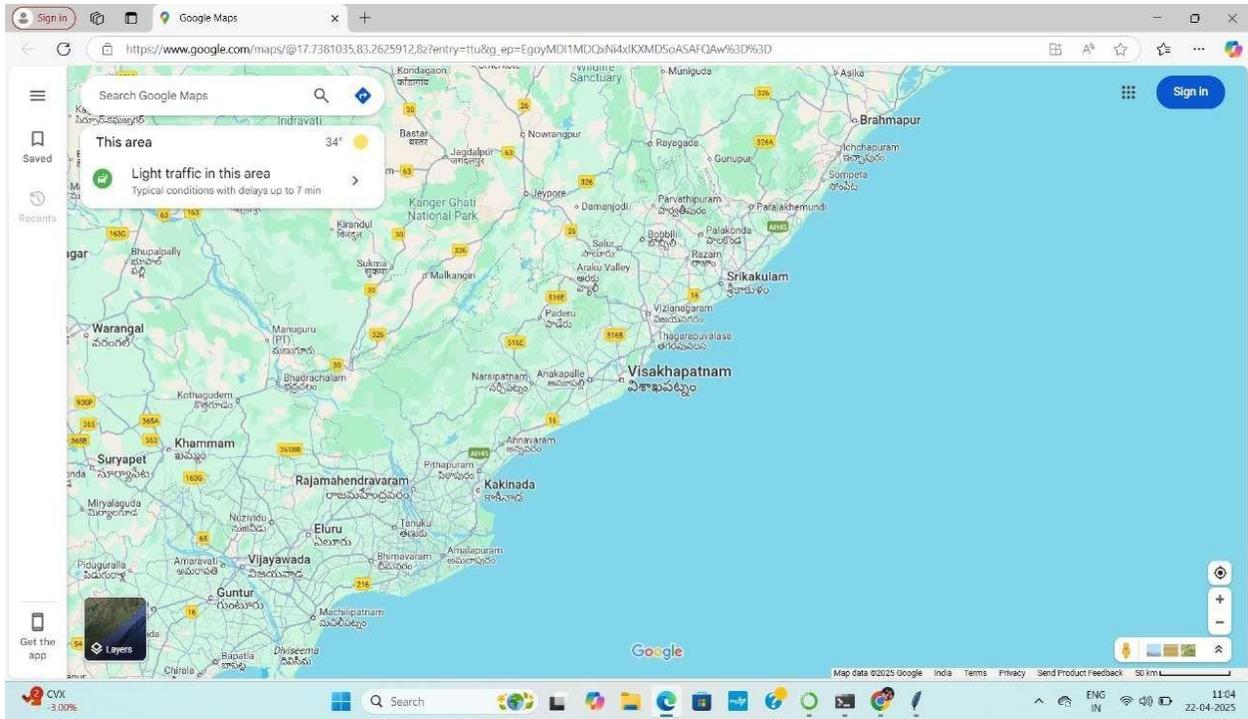Voice Input: open map
"Stop" Output:

Fig1: Google maps for light traffic in the area

11. Open Wikipedia
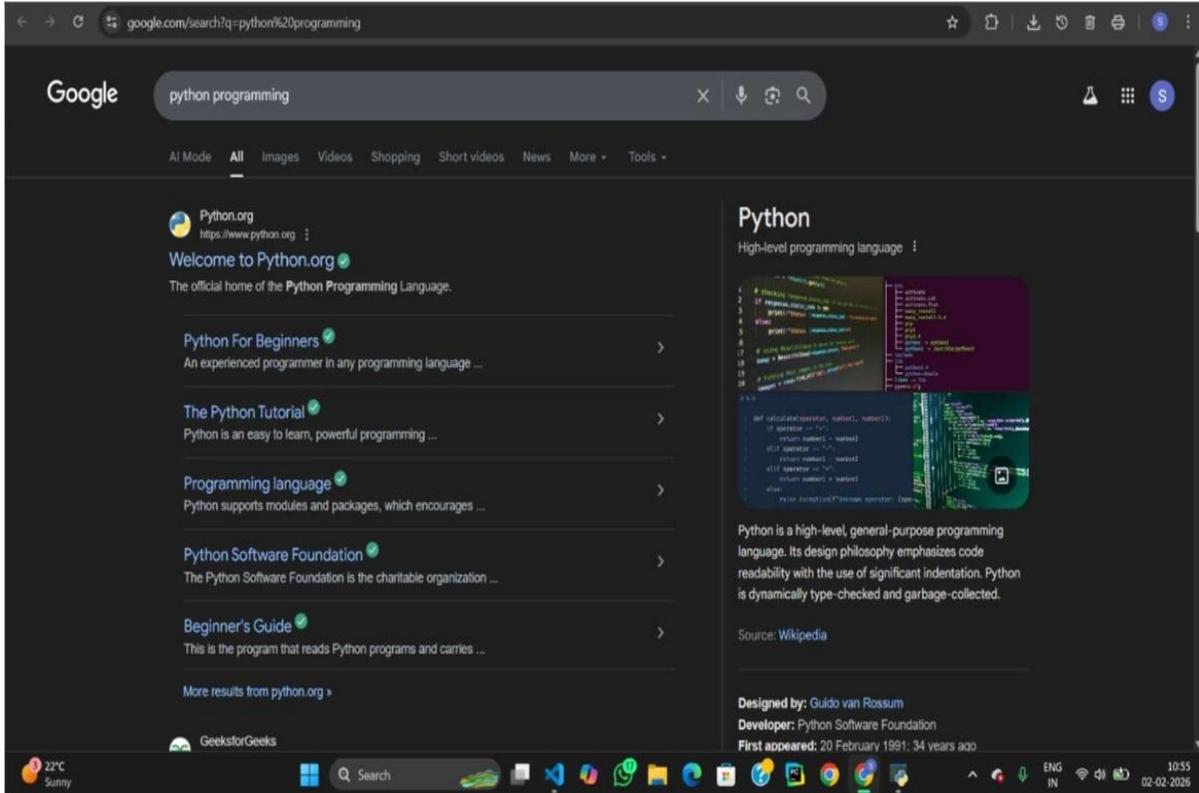Voice Input: python programming Output:

Fig 2: Python programming

12. Exit Command
Voice Input: "Stop"
Output:
Assistant says: "Have a good day sir!" "Please give your rating from 1 to 5."
Rating: "Thank you for giving your valuable me."
Program_closes

V. CONCLU0SION

As stated before, "voice assistant is one of the biggest problem solvers" and you can see that in the proposals with the examples that it is in fact one of the biggest problem solvers of the current world. We can see that voice assistant is one of the major evolving artificial intelligence in the current world once again on seeing the proposal examples because at the past, the best feature which a voice assistant had was telling the date and searching the web and giving the results but now look at the functions that it can do so with this, we can say that it is a evolving software in the current world. The main idea is to develop the assistant even more advanced than it is now and make it the best ai in the world which will save an ample of time for its users. I would like to conclude with the statement that we will try our best and give one of the best voice assistants which we are able to.

REFERENCES

[1] Kiesel, J., Bahrami, A., Stein, B., Anand, A., & Hagen, M. (2018, June). Toward voice query clarification. In The 41st international acm sigir conference on research & development in information retrieval (pp. 1257-1260).

[2] Godinez, J. E., & Jamil, H. M. (2019, April). Meet Cyrus: the query by voice mobile assistant for the tutoring and formative assessment of SQL learners. In Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing (pp. 24612468).

[3] S.S.D.K. Maha Lakshmi, Umamaheswararao Mogili, Sravya Eluri, Dogga Ramachandra Rao. (2023), "Online Dynamic Out Patient Queue System for Automated Token Generation in Hospitals", Science, Technology and Development Journal, Volume XII, Issue VII, pp 71-78, DOI: 23.18001.STD.2023.V12I07.23.37707.

[4] Sree, S. V. D. T., Mogili, U. M. R., & Ampoly, K. V. (2025) Enhancing Security in Wearable Computing: A Lightweight Authenticated Key Exchange Scheme, International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211, Volume 13, Issue 5, pp 3103-3108.

[5] Anjali, S., Mogili, U., & Ampolu, K. V. (2025) Efficient Key-Based Encryption and Authentication for Advanced Digital Forensic Storage Security, International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211, Volume 13, Issue 5, pp 3097-3102.

[6] Adithya, P. U., Mogili, U., & Mondru, J. T. (2025) A Novel Parity Authenticator-Based Zero-Knowledge Auditing Approach for Secure Cloud Data Management, International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211, Volume 13, Issue 5, pp 994-999.

[7] Kanakala Pranay Raj, Umamaheswararao Mogili. (2020), "Cloud-of-Cloud: A Novel Protocol for Secure Data Storage and Sharing in Multi-Cloud Environment", Journal of Interdisciplinary Cycle Research (JICR), Volume XII, Issue VI, pp 2201-2209, DOI:18.0002.JICR.2020.V12I6.008301.3 171227.

[8] Mogili, U., Mohamed, A., & Kasup, C. (2023, December). Mechanism of Data Sharing Using Secured Keyword Search in Cloud Computing. In Conference of Innovative Product Design and Intelligent Manufacturing System (pp. 483-494). Singapore: Springer Nature Singapore.

[9] Hassan Awadallah, A., Gurunath Kulkarni, R., Ozertem, U., & Jones, R. (2019, October). Characterizing and predicting voice query reformulation. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (pp. 543-552).

[10] Dambanemuya, H. K., & Diakopoulos, N. (2021). Auditing the information quality of news- related queries on the Alexa voice assistant. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), 1-21.

[11] Guy, I. (2021, July). Searching by talking: Analysis of voice queries on mobile web search. In Proceedings of the 39th International ACM

SIGIR conference on Research and Development in Information Retrieval (pp. 35-44).

[12] Sannigrahi, S., Fraga-Silva, T., Oualil, Y., & Van Gysel, C. (2023, July). Synthetic query generation using large language models for virtual assistants. In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 2837-2841).

[13] Van Gysel, C. (2023, July). Modeling spoken information queries for virtual assistants: Open problems, challenges and opportunities. In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 3335-3338).

[14] Sun, Z., Lu, S., Ma, C., Liu, X., & Guo, C. (2023). Query expansion and entity weighting for query reformulation retrieval in voice assistant systems. arXiv preprint arXiv:2202.13869.

[15] Hoy, M. B. (2024). Alexa, Siri, Cortana, and more: an introduction to voice assistants. Medical reference services quarterly, 37(1), 81-88.

[16] Altuwaiyan, T., Hadian, M., Rubel, S., & Liang, X. (2024, June). Exploiting privacy- preserving voice query in healthcare-based voice assistant system. In ICC 20202020 IEEE International Conference on Communications (ICC) (pp. 1-6). IEEE.

[17] Van Gysel, C. (2023, July). Modeling spoken information queries for virtual assistants: Open problems, challenges and opportunities. In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 3335-3338).

[18] Dambanemuya, H. K., & Diakopoulos, N. (2021). Auditing the information quality of news-related queries on the Alexa voice assistant. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), 1-21.

[19] Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: an introduction to voice assistants. Medical reference services quarterly, 37(1), 81-88.

[20] Ammari, T., Kaye, J., Tsai, J. Y., & Bentley, F. (2019). Music, search, and IoT: How people (really) use voice assistants. ACM Transactions on Computer Human Interaction (TOCHI), 26(3), 1-28.

[21] Mogili, U., Ampolu, K. V., Rajasekharam, B., & Timothy, M. J. AI-Driven Interaction in AR Environments, in Journal of Digital Economy, 2024, Volume 3, Issue 1, pp. 228-234.

[22] Timothy, M. J., Rajasekharam, B., Ampolu, K. V., & Mogili, U. Threat Detection Using AI in Cybersecurity Systems, in IJIS, 2023, Volume 7, Issue 1, pp. 1-7.

[23] Ampolu, K.V., Mogili, U., Timothy, M. J., & Rajasekharam, B. Machine Learning Models for Predictive Maintenance, in IJIS, 2022, Volume 6, Issue 4, pp. 1-7.

[24] Rajasekharam, B., Timothy, M. J., Mogili, U., Ampolu, K.V., Machine Learning Models for Predictive Maintenance, in JDE, 2023, Volume 2, Issue 2, pp. 95-101.

[25] Soujania, B., Ampolu, K. V., Timothy, M. J., & Mogili, U. (2025) Classifying Disease Information Forums through Semantic Similarity-Based Machine Learning, Science, Technology and Development Journal, Volume XIV, Issue II, pp 67-75.

[26] B Satish Kumar, Kavitha C., Mogili, U.R., S. Pallam Shetty (2022). "Application of Machine Learning To Enhance the Performance of The Prophet Routing Protocol For Delay Tolerant Networks". Journal for Basic Sciences, Volume 23, Issue 5, 2107-2116, DOI:10.37896/JBSV23.5/2278.

[27] I. Sree Geeta, Umamaheswararao Mogili. (2022), "Use of Several Machine Learning Algorithms for Effective Prediction of Cyberbullying", International Journal of Creative Research Thoughts, Volume 10, Issue 6, pp 17.

[28] Mogili, U., & Mohamed, A. (2023, November). Artificial intelligence and machine learning in the fields of education, medical, and smart phones. In AIP conference proceedings (Vol. 2917, No. 1, p. 050012). AIP Publishing LLC.